

RESEARCH

Open Access

Conditional downsampling for energy-efficient communications in wireless sensor networks

Joan Enric Barceló-Lladó*, Antoni Morell and Gonzalo Seco-Granados

Abstract

This paper deals with the power limitations in a wireless sensor network scenario. Concretely, we propose to use a conditional downsampling encoder (CDE) at the sensing nodes as an energy-efficient solution for the communication problem. It exploits the knowledge about the signal structure, which is assumed to be time-correlated, in order to decrease the sampling rate and hence to reduce the number of transmissions within the network. We analytically assess the performance of the CDE in terms of quadratic distortion, from which we derive closed-form expressions when it is combined with one of the two decoders: the step decoder and the predictive decoder. Moreover, we propose two methodologies to design the CDE in order to guarantee a given coding rate. We also compare the CDE, both analytically and experimentally, with other classical decimator techniques, which are the deterministic downsampling encoder and the probabilistic downsampling encoder. Numerical simulation validates our analytical results. Moreover, we compare the obtained quadratic distortion and extract the conclusions of the capabilities of the studied encoding-decoding schemes.

1 Introduction

1.1 Motivation and previous work

Wireless sensor network (WSN) design is currently one of the most challenging topics in the communications field. In particular, WSNs are severely energy-constrained because they consist of many small, cheap, and power-limited nodes, whose batteries cannot be recharged in most cases. Hence, the application of energy-efficient algorithms turns out to be crucial.

Following with this motivation, many energy-efficient strategies can be found in order to mitigate the energy costs and hence increase the lifetime of the WSN. Without the aim of being exhaustive, we point out some examples:

- Energy-aware routing for cooperative WSNs and *ad hoc* networks [1,2]. These techniques seek the optimum path that minimizes the total spent energy in multihop WSNs.
- Signal processing techniques for minimum-power distributed transmission schemes [3,4]. Using distributed beamforming techniques, the nodes can

decrease the transmitted power at the same time that they increase the total throughput of the network.

- Data-aware techniques to reduce energy by efficient information processing [5,6]. By means of signal processing techniques, the network exploits the inherent structure and properties of the measured signal in order to sample the data and therefore reduce the associated energy costs.

Our study falls in the third category and may be complementary to the other approaches. Actually, we propose to encode the sensed data removing redundancy in the time domain. Many transmission schemes use non-causal transmissions such as block coding. In these cases, the source collects a number of contiguous time samples in order to compress them by removing part of (or all) the redundancy among them. Within this group of encoders-decoders, a large amount of different techniques can be found. Albeit these transmissions are very appropriate for high-rate transmissions and/or delay-tolerant communications, these non-causal transmission schemes may not be applicable in some scenarios because block transmissions are not always allowed due to delay constraints and/or low symbol rates of the source.

For delay-sensitive applications such as real-time monitoring in WSNs, where the reconstruction of the signal

*Correspondence: joanenric.barcelo@uab.es

Department of Telecommunications and Systems Engineering, Universitat Autònoma de Barcelona, Bellaterra, Barcelona 08193, Spain

must take place at the same time instant as the corresponding input measurement, causal source codes are more convenient. Hence, a source code is said to be causal if the n th decoded sample depends on the output signal only through its first n components or, in other words, depends on the past and present outputs but not on future ones. Quantizers, delta modulators, differential pulse code modulators, and adaptive versions of these are all causal in the above sense. The basic properties of causal source codes have been introduced in 1982 in [7], and related works have been expanded so far. The work in [8] extends the general results of [7] for the case where *side information*, i.e., extra information that is correlated with the source, is available at the encoder and the decoder. In addition, a causal source code is called a zero-delay or sequential code if both the encoder and the decoder are causal (note that for the causal source code definition, the assumption of causality is only at the decoder) [9,10].

In the literature, there are several zero-delay coding systems. One of the most common zero-delay coding systems is the well-known *differential pulse code modulation* (DPCM). In a nutshell, the current sample to be coded is predicted from previously coded samples. This prediction is used as a reference, and it is compared with the current sample. Hence, the output of the encoder is the prediction error. The inverse operation takes place at the decoder side. According to [11], DPCM was first introduced in a US patent by C. Cutler in 1952. Since then, many results have appeared. In particular, the *autoregressive* (AR) model has received special attention for the study of zero-delay coding schemes. Some of the early works on AR models date back to the 1960s. The works in [12,13] analyze the quadratic rate distortion of DPCM (the reader can find an extended description of the rate distortion in Chapter 13 of [14]). The work in [15] extends these results assuming a Gaussian distribution of the predicted error. Other works proposed algorithms for non-uniform quantizers optimized in order to minimize the distortion rate [16].

Later works, as the one in [17], try to particularize the results obtained by the DPCM also for the case of low bit rates. In such cases, the system performance becomes worse. Then, the classic DPCM encoder is modified in order to achieve better performance in terms of rate distortion for a low-bit-rate regime.

Recent works on this field have tried to unify the theoretical limits of the DPCM (and other zero-delay schemes) for AR models with other information theory concepts. The authors in [11] provide analytical results for the existing duality between the rate distortion of an AR process with the capacity of the inter-symbol interference channels. By contrast, other works such as [9] follow an information theoretical approach that adjusts the upper and lower limits of the

rate distortion for generic zero-delay schemes using the mutual information as a measure of the achievable rate.

1.2 Our contribution

Our proposed work also follows the same sequential transmission approach exposed above. Concretely, the approach of this paper is similar to that of [6], where the authors seek for the optimal sampling in a WSN scenario with correlated sources. However, we present the problem from a more realistic energy-efficient perspective. According to the results in the literature about energy consumption in sensor networks [18], the main source of energy spent in a sensor is the power dedicated to maintain the sensor awake. Concretely, most energy is consumed by the elements of the front-end [19]. Therefore, our goal is to reduce the number of total transmissions in order to keep the sensors in sleep mode as long as possible.

Note that for the complete characterization of the performance of real communication systems, several metrics should be evaluated, e.g., the robustness against noise in terms of the signal-to-noise ratio, the quantization error as a function of the codification scheme, or the bit error rate related to a selected modulation. However, in this paper, we only focus on the study of the downsampling distortion (see Section 2.2) as a figure of merit of the quadratic reconstruction error introduced by a downsampling technique at the fusion center. The study of other performance metrics, although interesting, is out of the scope of this paper.

In particular, we study downsampling techniques in which the samples of an input signal are either blocked or transmitted following a given criterion. For that purpose, we propose a downsampling encoding scheme called conditional downsampling encoder (CDE). A CDE benefits from the existing time correlation in the measured signal in order to sequentially elaborate the decimator pattern. Typically, the readings in WSNs are space-time-correlated, and hence, strategies in the two domains can potentially improve the accuracy of the signal recovered at the receiver side. However, note that considering not only the time correlation but also the space correlation at the sensing nodes would require intensive inter-node communication. Since this approach would penalize in terms of signaling, complexity, and energy consumption, we have discarded it. Basically, the CDE predicts the current sample using a linear estimation and takes this prediction as a reference. Then, the transmission is blocked if the prediction error does not exceed a given threshold and transmitted otherwise. It is clear that a key step of the CDE design is to determine the threshold that ensures a sample rate reduction of a factor γ . Therefore, two different threshold designs are proposed in this paper.

Clearly, the CDE presents some similarities with the DPCM in the sense that both schemes use (linear) prediction as a reference in order to encode the input signal. However, they present important differences as well, which can be summarized as follows:

- A DPCM produces an outcome sample for each input sample. In other words, it does not change the sample rate. On the contrary, the CDE (and also the *deterministic downsampling encoder* (DDE) and the *probabilistic downsampling encoder* (PDE)) reduces the sample rate. This behavior is very convenient in some energy-constrained scenarios, such as WSNs, since the total number of transmissions is reduced by a factor γ , increasing the energy efficiency of the network.
- While a DPCM works at the symbol level, the CDE does at the sample level. Thus, the downsampling encoder-decoder schemes studied in this paper are not exclusive to the DPCM or other zero-delay coding techniques. Actually, they can be used on top of them when the signal is transmitted.

In addition, we compare the performance loss of CDE with different encoding-decoding pairs when the number of samples is reduced by a factor γ . In particular, we study the following two downsampling criteria: (1) a DDE and (2) a PDE.

A DDE works as a decimator, i.e., it reduces the number of samples following a deterministic pattern. Hence, the DDE selects only one in γ^{-1} samples, where γ^{-1} is typically a natural number.

A PDE slightly differs from a common decimator since it reduces the number of samples following a probabilistic pattern, i.e., one sample will be transmitted with probability γ and otherwise blocked with probability $1 - \gamma$. This method eliminates the restriction of γ^{-1} to be a natural number. However, we analytically show that a DDE outperforms a PDE in terms of quadratic distortion.

On the other hand, the decoder at the fusion center recovers the original sampling rate by upsampling the signal. We study two possible decoders: (1) a *step decoder* (SD) and (2) a *predictive decoder* (PD). A SD reconstructs the missing samples by replicating the last decoded sample. This does not require any side information knowledge. On the contrary, the PD reconstructs the missing samples by linear prediction (as in the CDE case). We analytically show the improvements in terms of quadratic distortion when the samples are predicted rather than simply replicated.

Hence, we give analytical expressions for the quadratic distortion of the following downsampling encoding-decoding pairs: DDE-SD, DDE-PD, PDE-SD, and PDE-PD. Furthermore, we also provide accurate approximations

for the quadratic distortion of CDE-SD and CDE-PD. Numerical simulations support our proposed analytical expressions.

1.3 Organization of the paper

The rest of the paper is organized as follows: In Section 2, we introduce the assumptions and the scenario considered throughout the paper. Section 3 presents the proposed CDE as well as the other encoding-decoding schemes under study. The analytical expressions of the downsampling distortion for the proposed CDE are detailed in Section 4. Also, two different design strategies are presented in this section. The analytical expressions of the downsampling distortion for other encoding-decoding schemes are detailed in Section 5. Simulation results are shown in Section 6. Conclusions and suggestions for future research are drawn in Section 7.

2 System model and assumptions

Let us consider a WSN configured in star topology that monitors a given physical scalar magnitude such as temperature or humidity. The network is composed of two types of nodes: (1) a set of S sensing nodes that transmit wirelessly the measurements to (2) one fusion center that manages, gathers, and processes the measurements from the sensing nodes.

2.1 Assumptions on the signal model

We consider the signal modeled as an S -dimensional stochastic process, namely^a,

$$\mathbf{X} = [\mathbf{x}(1) \mathbf{x}(2) \dots \mathbf{x}(N)], \quad (1)$$

where $\mathbf{x}(n) = [x_1(n) x_2(n) \dots x_S(n)]^T$ and $x_s(n)$ denotes the measurement of the s th sensor at the sample time n and N denotes the number of time samples in the observation window. Let $x_s(n)$ be a real and time-discrete autoregressive model of order 1 (AR-1), variance σ_x^2 , and sampled at a rate \mathcal{R} , which is commonly assumed in the signal processing literature in order to model real sources [20]. It is defined as

$$x_s(n) = \rho x_s(n-1) + z(n), \quad \text{for } n = 1, 2, \dots \quad (2)$$

The autoregression coefficient is denoted by $\rho \in [0, 1]$ and assumed to be constant during the transmission. The random process $z(n)$ is a sequence of Gaussian-distributed and independent random variables with zero mean and variance σ_z^2 .

Without loss of generality, we also assume that the variance of the measurement $x_s(n)$, i.e., σ_x^2 , is equal to 1. Therefore, the variance of the noise is well known, and it is $\sigma_z^2 = 1 - \rho^2$.

2.2 Assumptions on the system model

We do not assume any coordination among sensing nodes. Hence, each one will act non-cooperatively. It reduces the required signaling in comparison to cooperative communications and also allows us to focus our analysis only in the communications between one sensing node and the fusion center without loss of generality. The transmission model under consideration is a generic one, and it is illustrated in Figure 1.

Note that for simplicity, we have replaced the notation $x_s(n)$ by $x(n)$. Furthermore, we require that the signal $x(n)$ is transmitted in a zero-delay manner from the source to the destination. Throughout this paper, we understand for zero-delay transmission when for each sample at time n , the receiver will have a reconstruction of the signal $x(n)$. Furthermore, for time instant n , we are not interested in $x(n-1)$ anymore, so delay-tolerant strategies (such as block encoding schemes) are not feasible. Following this constraint, we will look for encoders that allow us to reduce the sample rate sample-by-sample in real time.

Hence, we consider a non-linear encoder with a coding rate γ at the sensing nodes. In our particular case, the encoder selects which samples from $x(n)$ are going to be transmitted with a rate of γ , and the rest will be discarded. The selected samples are represented by $y(n)$; therefore, note that $y(n)$ is only defined for those time slots in which the encoder decides to transmit.

Moreover, we consider non-linear decoders in order to recover an approximation of $x(n)$, i.e., $\tilde{x}(n)$, from $y(n)$ at the fusion center. Roughly speaking, the decoder will construct $\tilde{x}(n)$ copying the samples of $y(n)$ when the transmission exists and predicting the rest otherwise.

Definition 1. For a given pair of encoder-decoder, the sink will receive $\tilde{x}(n)$ with a given downsampling distortion. It defines the quadratic distortion introduced by the given downsampling encoder-decoder pair e - d as

$$\mathcal{D}(e, d) = \mathbb{E}[(x(n) - \tilde{x}(n))^2]. \quad (3)$$

3 Downsampling transmission schemes

3.1 Different encoding alternatives

We compare our proposed CDE with two selected downsampling encoders among many other possibilities. These are (1) the DDE and (2) the PDE. They have been chosen since they are simple and because many other strategies can be derived from them.

In order to describe the selected encoders, we first need to introduce the following definition:

Definition 2. The transmission support function of an encoder e , named $g_e(n)$, is an indicator function which takes the value 1 when the transmission exists and 0 otherwise.

3.1.1 Deterministic downsampling encoder

This encoder is the simplest and acts as a typical decimator. Its transmission support function is

$$g_{\text{DDE}}(n) = \begin{cases} 1 & \text{when } n \bmod \gamma^{-1} = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Note that for uniform downsampling, the DDE is only defined for compression rates γ of the form $\gamma^{-1} \in \mathbb{N}$.

3.1.2 Probabilistic downsampling encoder

This encoder solves the limitation of DDE that γ^{-1} is a natural number. Basically, the symbol $x(n)$ will be transmitted following a given probabilistic pattern. Thus, the transmission support function is

$$g_{\text{PDE}}(n) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p. \end{cases} \quad (5)$$

It is straightforward to see that in order to guarantee a compression rate of γ , the value of the transmission probability p should be $p = \gamma$.

3.1.3 Conditional downsampling encoder

Previous encoders do not assume any memory or prior information of the signal of interest $x(n)$. On the contrary, the CDE uses the available information in order to decide whether the signal should be transmitted or not. In particular, we analyze the cases where the available information is either the last decoded sample $\tilde{x}(n-1)$ or a linear prediction using the linear Wiener filter (LWF) solution in [21] with a given observation vector $\tilde{\mathbf{x}}(n)$. The available information is compared with the signal of interest $x(n)$. If the absolute value of the difference is higher than a given threshold Δ , the encoder will transmit the signal. Otherwise, if the difference is below Δ , the transmission is blocked. Mathematically, for the first case,

$$g_{\text{CDE}}(n) = \begin{cases} 1 & \text{if } |x(n) - \tilde{x}(n-1)| > \Delta \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

For the LWF prediction, the CDE is

$$g_{\text{CDE}}(n) = \begin{cases} 1 & \text{if } |x(n) - \hat{x}(n)| > \Delta \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Although this scheme is quite simple, it has two main complications: (1) the LWF predictor assumes the knowledge of the correlation parameters \mathbf{R} and \mathbf{r} or at least good estimates of them, and (2) the threshold Δ should be designed in such a way that it ensures a coding rate of γ . The first problem adds some complexity to the system but can be efficiently solved using existing correlation estimators [22]. The second one is addressed later in Section 4.

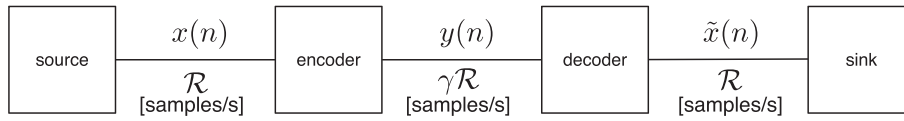


Figure 1 Block diagram of a generic encoder-decoder transmission scheme. The source is generating the desired signal $x(n)$ at a rate R (samples/s). The signal $y(n)$ is the encoded version of $x(n)$, and it is transmitted at a rate γR , where $\gamma \leq 1$. The signal $\tilde{x}(n)$ is the reconstruction of the desired signal $x(n)$ after the decoder.

3.2 Different decoding alternatives

As for the encoding strategies, we select two decoders from a bunch of possible solutions. The first one is probably the simplest and does not require any knowledge of the correlation parameters, while the second one exploits the signal correlation in order to achieve higher prediction accuracy.

3.2.1 Step decoder

It is the simplest decoder. It just copies the value of $y(n)$ into $\tilde{x}(n)$ when $g_e(n) = 1$ or maintains the last decoded value $\tilde{x}(n-1)$ if $g_e(n) = 0$. The decoder function is described as

$$d_{SD}(n) = \begin{cases} \tilde{x}(n) = y(n) & \text{if } g_e(n) = 1 \\ \tilde{x}(n) = \tilde{x}(n-1) & \text{otherwise.} \end{cases} \quad (8)$$

This approach is very typical when the source is sensing a given time-correlated phenomenon. Since it is assumed to be slow changing, the magnitude is maintained until we receive an update.

3.2.2 Predictive decoder

If we take advantage of the time correlation properties of $x(n)$, we can obtain lower downsampling distortion than for the SD case. The behavior is similar to the previous decoder SD, but in this case, when $g_e(n) = 0$, the PD predicts $x(n)$ using LWF instead of replicating $\tilde{x}(n)$. Mathematically,

$$d_{PD}(n) = \begin{cases} \tilde{x}(n) = y(n) & \text{if } g_e(n) = 1 \\ \tilde{x}(n) = \hat{x}(n) & \text{otherwise.} \end{cases} \quad (9)$$

4 Downsampling distortion of the conditional downsampling encoder

4.1 Signal prediction using incomplete observation vectors

Let the observation vector $\tilde{\mathbf{x}}(n) \in \mathbb{R}^N$, where $\tilde{\mathbf{x}}(n) = [\tilde{x}(n-1) \tilde{x}(n-1) \dots \tilde{x}(n-N)]^T$, be an incomplete version of $\mathbf{x}(n)$. The vector $\tilde{\mathbf{x}}(n)$ is constructed using the N last decoded samples. This is because the decoder does not necessarily know all the values of $\mathbf{x}(n)$ and only knows the decoded ones. Hence, some values of $\tilde{\mathbf{x}}(n)$ are replicas of $\mathbf{x}(n)$, and the rest are predicted values $\hat{x}(n)$.

Definition 3. Let the vector $\tilde{\mathbf{x}}_t$ be an instance of $\tilde{\mathbf{x}}(n)$ where the last true sample was received at time $n-t$. Mathematically,

$$[\tilde{\mathbf{x}}_t(n)]_j = \begin{cases} \hat{x}(n-j) & \text{if } j < t \\ x(n-j) & \text{if } j = t. \end{cases} \quad (10)$$

Theorem 1. If $\tilde{\mathbf{x}}_t(n)$ is used as the observation vector of the LWF, the mean square error (MSE) is degraded as

$$\text{MSE}_t = 1 - \rho^{2t}. \quad (11)$$

Proof. It is proved by induction. First let us assume the case where the vector $\tilde{\mathbf{x}}_2(n)$ is of the form $\tilde{\mathbf{x}}_2(n) = [\hat{x}(n-1) x(n-2) \dots x(n-N)]^T$, that is, all the positions in the vector correspond to true measurements except for the first one. In this case,

$$\begin{aligned} \mathbb{E}[|x(n) - \mathbf{w}^H \tilde{\mathbf{x}}_2(n)|^2] &= \mathbb{E}[|x(n) - \rho \hat{x}(n-1)|^2] \\ &= \mathbb{E}[|x(n) - \rho \mathbf{w}^H \tilde{\mathbf{x}}_1(n-1)|^2] \\ &= 1 - 2\rho^2 \mathbb{E}[x(n)x(n-2)] \\ &\quad + \rho^4 \mathbb{E}[x(n-2)x(n-2)] = 1 - \rho^4. \end{aligned} \quad (12)$$

For the case where $\tilde{\mathbf{x}}_3(n)$ is of the form $\tilde{\mathbf{x}}_3(n) = [\hat{x}(n-1) \hat{x}(n-2) x(n-3) \dots x(n-N)]^T$, the MSE is degraded as

$$\begin{aligned} \mathbb{E}[|x(n) - \mathbf{w}^H \tilde{\mathbf{x}}_3(n)|^2] &= \mathbb{E}[|x(n) - \rho \hat{x}(n-1)|^2] \\ &= \mathbb{E}[|x(n) - \rho \mathbf{w}^H \tilde{\mathbf{x}}_2(n-1)|^2] \\ &= \mathbb{E}[|x(n) - \rho^2 \mathbf{w}^H \tilde{\mathbf{x}}_1(n-2)|^2] \\ &= 1 - 2\rho^3 \mathbb{E}[x(n)x(n-3)] \\ &\quad + \rho^6 \mathbb{E}[x(n-3)x(n-3)] \\ &= 1 - 2\rho^6 + \rho^6 = 1 - \rho^6. \end{aligned} \quad (13)$$

It is straightforward to conclude that for the general case where $\tilde{\mathbf{x}}_t(n)$ is of the form $\tilde{\mathbf{x}}_t(n) = [\hat{x}(n-1) \dots \hat{x}(n-t+1) x(n-t) \dots x(n-N)]^T$, the MSE is degraded as

$$\mathbb{E}[|x(n) - \mathbf{w}^H \tilde{\mathbf{x}}_t(n)|^2] = 1 - \rho^{2t}. \quad (14)$$

□

Corollary 1. For a given ρ , the MSE is only a function of the position of the last true measurement in the observation vector for an AR-1 process. Furthermore, it is not dependent on the dimension N of $\tilde{\mathbf{x}}_t(n)$.

Proof. The proof of the first statement is straightforward, and it is enough to verify that the MSE obtained by $\tilde{\mathbf{x}}_t(n)$ and $\tilde{\mathbf{x}}'_t(n)$, where

$$\tilde{\mathbf{x}}'_t(n) = [\hat{x}(n-1) \dots x(n-t) \dots \hat{x}(n-N)]^T, \quad (15)$$

is the same. Then, let us consider, for example, $t = 2$,

$$\begin{aligned} \mathbb{E}[|x(n) - \mathbf{w}^H \tilde{\mathbf{x}}'_2(n)|^2] &= \mathbb{E}[|x(n) - \rho \hat{x}(n-1)|^2] \\ &= \mathbb{E}[|x(n) - \mathbf{w}^H \tilde{\mathbf{x}}_2(n)|^2] = 1 - \rho^4. \end{aligned} \quad (16)$$

Moreover, for observation vectors that only contain estimated measures (i.e., $t > N$), the MSE also follows (11). It can be observed that if $t = N + 1$, then the MSE is

$$\begin{aligned} \mathbb{E}[|x(n) - \mathbf{w}^H \tilde{\mathbf{x}}_{N+1}(n)|^2] &= \mathbb{E}[|x(n) - \rho^T \hat{x}(n-N)|^2] \\ &= \mathbb{E}[|x(n) - \rho^N \mathbf{w}^H \tilde{\mathbf{x}}_1(n-N)|^2] \\ &= \mathbb{E}[|x(n) - \rho^{N+1} x(n-N-1)|^2] \\ &= 1 - \rho^{2(N+1)}. \end{aligned} \quad (17)$$

□

Similarly, if the last transmitted sample $x(n-t)$ is directly used as a reference or prediction, the MSE when the observation vector is $\tilde{\mathbf{x}}_t$ is degraded as

$$\begin{aligned} \mathbb{E}[(x(n) - x(n-t))^2] &= \mathbb{E}[x^2(n)] - 2\mathbb{E}[x(n)x(n-t)] \\ &\quad + \mathbb{E}[x^2(n-t)] \\ &= 1 - 2\rho^t + 1 = 2(1 - \rho^t). \end{aligned} \quad (18)$$

Hence, the probability that the last true sample of the vector $\tilde{\mathbf{x}}(n)$ is in the position t depends directly on the downsampling criteria used at the encoder. Therefore, in order to compute the downsampling distortion for the CDE, we need to compute the probability of occurrence of the event t , or what is the same, the probability that the observation vector $\tilde{\mathbf{x}}$ is actually $\tilde{\mathbf{x}}_t$. Next, we illustrate the CDE problem using a Markov chain (MC) model.

4.2 The Markov chain solution for the incomplete observation vector case

Let a MC model be a discrete time process where a random variable $E(n)$ is changing in time. The MCs have the property that to be in a state t , i.e., $E(n) = t$, only depends on the previous state, i.e., $E(n-1)$. This property is very interesting in order to model AR-1 processes. Moreover, a MC is said to be homogeneous when the probability of transition between the states of $E(n)$ is invariant in time, i.e.,

$$p_{ij} = P(E(n)=j|E(n-1)=i) \in [0, 1] \text{ where } i, j = 0, 1, \dots, T-1. \quad (19)$$

Definition 4. Let the matrix $\mathbf{T} \in \mathbb{R}^{T \times T}$ denote the transition matrix of a homogeneous MC process of T states where

$$[\mathbf{T}]_{ij} = p_{ij} \text{ for } i, j = 1, \dots, T-1 \quad (20)$$

and each row represents a probability distribution, so $[\mathbf{T}^T]_i \mathbf{1} = 1$.

Definition 5. Let the vector $\mathbf{p} \in \mathbb{R}^T$ denote the stationary probability vector of a homogeneous MC process of T states and any vector that holds the stationary conditions

$$\mathbf{p}^T = \mathbf{p}^T \mathbf{T} \text{ and } \mathbf{p}^T \mathbf{1} = 1 \quad (21)$$

where $\mathbf{p} = [P_0 P_1 \dots P_{T-1}]^T$ contains the probabilities to be in each state $t = 0, 1, \dots, T$ in the stationary regime of the MC process.

4.3 The Markov chain model for the CDE

In this section, we analytically evaluate the performance of the proposed CDE with both PD and SD decoders in terms of the downsampling distortion.

The CDE can be modeled following the infinite Markov chain in Figure 2. The state $E(n) = 0$ means that in time n , the transmission exists. Similarly, the state $E(n) = t$, for $t \neq 0$, means that the sample $n-t$ was the last to be transmitted. The transition matrix (with dimension $T \rightarrow \infty$) that describes the process of the CDE is

$$\mathbf{T}_{\text{CDE}} = \begin{bmatrix} p_{0,0} & p_{0,1} & 0 & \dots \\ p_{1,0} & 0 & p_{1,2} & \\ \vdots & \vdots & & \ddots \end{bmatrix}. \quad (22)$$

From the stationary condition in (21), we can obtain the following relations:

$$P_t = p_{t-1,t} P_{t-1}; \text{ thus, } P_t = P_0 \prod_{i=1}^t p_{i-1,i}, \quad (23)$$

where by definition $\sum_{i=1}^{\infty} P_i = 1 - P_0$. Moreover, after some algebraic manipulations,

$$\frac{1 - P_0}{P_0} = \sum_{t=1}^{\infty} \left(\prod_{j=1}^t p_{j-1,j} \right). \quad (24)$$

It is easy to observe that there are infinite solutions for the transition probabilities p_{ij} . Thus, we address the design and the corresponding performance in the following sections.

4.4 Approximations for the downsampling distortion of the CDE-PD and CDE-SD

Following the scheme in (7), our aim is to design the threshold value Δ in order to guarantee that the source only transmits a fraction γ of the total samples. For the general case, we may have different values of Δ according

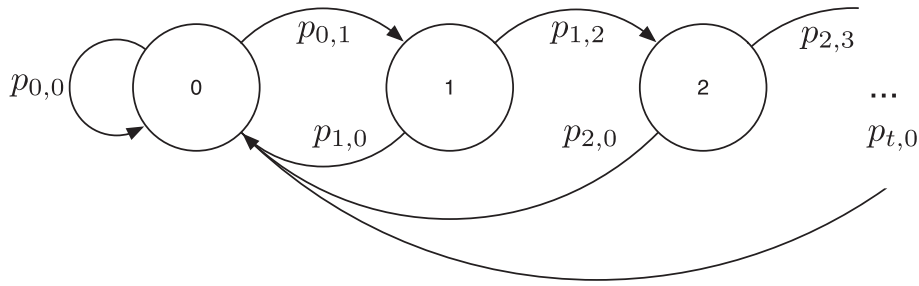


Figure 2 Infinite Markov chain that models the encoder CDE.

to each state t of the MC. Therefore, we define the threshold Δ_t as the threshold value applied to the state t .

The condition in (7) modifies the probability density function (pdf) of the error.

Definition 6. Let the conditional pdf $f(x| |x| < \Delta_t)$ be the pdf of x conditioned to $|x| < \Delta_t$. Mathematically,

$$f(x| |x| < \Delta_t) = \beta(\Delta_t)^{-1} f(x) \Pi\left(\frac{x}{2\Delta_t}\right), \quad (25)$$

where $f(x)$ is the original pdf of x and $\beta(\Delta_t) \in (0, 1)$ is

$$\beta(\Delta_t) = \int_{-\Delta_t}^{\Delta_t} f(x) dx. \quad (26)$$

Moreover, the *rectangular function* $\Pi(x)$ is defined as follows: $\Pi(x) = 0$ if $|x| > 0.5$, $\Pi(x) = 1$ if $|x| < 0.5$, and $\Pi(x) = 0.5$ if $|x| = 0.5$. This definition is summarized in Figure 3.

Lemma 1. Let $x \sim \mathcal{N}(0, \sigma^2)$. Then, the variance of the conditional pdf $f(x| |x| < \Delta_t)$ is

$$\text{var}(x| |x| < \Delta_t) = \frac{2}{\sqrt{2\pi\sigma^2}} \left(-\Delta_t \sigma^2 e^{\frac{-\Delta_t^2}{2\sigma^2}} + \frac{1}{2} \sqrt{2\pi\sigma^6} \text{erf}\left(\frac{\Delta_t}{\sqrt{2\sigma^2}}\right) \right). \quad (27)$$

Proof. Let x' define the random variable

$$x' \sim \{x_1 | |x| < \Delta_t\}, \quad (28)$$

where $x_1 \sim \mathcal{N}(0, \sigma^2)$. Hence,

$$\text{var}(x') = \text{var}(x| |x| < \Delta_t) = \int_{-\Delta_t}^{\Delta_t} x^2 f(x| |x| < \Delta_t) dx. \quad (29)$$

Using the relation

$$f(A|B) = \frac{f(A, B)}{P(B)}, \quad (30)$$

we obtain

$$\text{var}(x') = \int_{-\Delta_t}^{\Delta_t} x^2 \frac{f(x, |x| < \Delta_t)}{P\{|x| < \Delta_t\}} dx. \quad (31)$$

The term $P\{|x| < \Delta_t\}$ in the denominator is

$$P\{|x| < \Delta_t\} = \int_{-\Delta_t}^{\Delta_t} f(x) dx = \beta(\Delta_t). \quad (32)$$

So,

$$\text{var}(x') = \beta^{-1}(\Delta_t) \int_{-\Delta_t}^{\Delta_t} x^2 f(x, |x| < \Delta_t) dx. \quad (33)$$

Applying the same relation as that in (30), we obtain

$$\text{var}(x') = \beta^{-1}(\Delta_t) \int_{-\Delta_t}^{\Delta_t} x^2 f(x) P\{|x| < \Delta_t | x\} dx, \quad (34)$$

where the term $P\{|x| < \Delta_t | x\}$ is

$$P\{|x| < \Delta_t | x\} = \Pi\left(\frac{x}{2\Delta_t}\right). \quad (35)$$

Thus,

$$\begin{aligned} \text{var}(x') &= \beta^{-1}(\Delta_t) \int_{-\Delta_t}^{\Delta_t} x^2 f(x) dx \\ &= \frac{2}{\beta(\Delta_t) \sqrt{2\pi\sigma^2}} \left(-\Delta_t \sigma^2 e^{\frac{-\Delta_t^2}{2\sigma^2}} + \frac{1}{2} \sqrt{2\pi\sigma^6} \text{erf}\left(\frac{\Delta_t}{\sqrt{2\sigma^2}}\right) \right) \end{aligned} \quad (36)$$

that comes from the relation

$$\int_0^\epsilon x^2 e^{-\alpha x^2} dx = -\frac{\epsilon}{2\alpha} e^{-\alpha \epsilon^2} + \frac{1}{4} \sqrt{\frac{\pi}{\alpha^3}} \text{erf}(\epsilon \sqrt{\alpha}). \quad (37)$$

□

Definition 7. We define the conditional function $h(\sigma^2 | \Delta_t) : \mathbb{R} \rightarrow \mathbb{R}$ as

$$h(\sigma^2 | \Delta_t) = \text{var}(x| |x| < \Delta_t). \quad (38)$$

4.4.1 The pair CDE-PD

The knowledge of some prior information about the signal can notably reduce the MSE at the decoder compared to other classical methods. This is because only the samples with lower MSE are predicted, i.e., the ones that satisfy $|x(n) - \mathbf{w}^T \tilde{\mathbf{x}}_t(n)| < \Delta_t$, since they introduce less noise power at the decoder.

Lemma 2. Let $\text{MSE}_t^{\text{CDE-PD}}$ be defined as the mean square error when the observation vector is $\tilde{\mathbf{x}}_t(n)$. Then, the $\text{MSE}_t^{\text{CDE-PD}}$ is an approximation of $\text{MSE}_t^{\text{CDE-PD}}$ (i.e., the error introduced by the CDE-PD pair at the state t) and defined as

$$\text{MSE}_t^{\text{CDE-PD}} = h \left(1 - \rho^2 + \rho^2 \text{MSE}_{t-1}^{\text{CDE-PD}} | \Delta_t \right) \simeq \text{MSE}_t^{\text{CDE-PD}}. \quad (39)$$

Proof. For $t = 1$, the error $\text{MSE}_1^{\text{CDE-PD}}$ follows the conditional variance^b such that

$$\begin{aligned} \text{MSE}_1^{\text{CDE-PD}} &= \mathbb{E} \left[(x(n) - \mathbf{w}^T \tilde{\mathbf{x}}_1(n))^2 \mid |x(n) - \mathbf{w}^T \tilde{\mathbf{x}}_1(n)| < \Delta_1 \right] \\ &= \mathbb{E} \left[(\rho x(n-1) + z(n) - \rho x(n-1))^2 \mid \rho x(n-1) \right. \\ &\quad \left. + z(n) - \rho x(n-1) < \Delta_1 \right] \\ &= \mathbb{E} \left[z(n)^2 \mid |z(n)| < \Delta_1 \right] \\ &= \int_{-\Delta_1}^{\Delta_1} z(n)^2 f(z(n)) \mid |z(n)| < \Delta_1 dz(n). \end{aligned} \quad (40)$$

Using Definition 7 and since $z(n) \sim \mathcal{N}(0, \sigma_z^2)$ where $\sigma_z^2 = 1 - \rho^2$, the $\text{MSE}_1^{\text{CDE-PD}}$ is

$$\text{MSE}_1^{\text{CDE-PD}} = h \left(1 - \rho^2 | \Delta_1 \right). \quad (41)$$

For $t = 2$, the available knowledge is twofold: (1) we know that $|x(n) - \mathbf{w}^T \tilde{\mathbf{x}}_2(n)| < \Delta_2$, and (2) we also know

that in $t = 1$ the error was $|z(n-1)| < \Delta_1$. Therefore, the $\text{MSE}_2^{\text{CDE-PD}}$ can be written as

$$\begin{aligned} \text{MSE}_2^{\text{CDE-PD}} &= \mathbb{E} \left[(x(n) - \mathbf{w}^T \tilde{\mathbf{x}}_2(n))^2 \mid |x(n) \right. \\ &\quad \left. - \mathbf{w}^T \tilde{\mathbf{x}}_2(n)| < \Delta_2, |z(n-1)| < \Delta_1 \right], \\ &= \mathbb{E} \left[(\rho x(n-1) + z(n) - \rho \mathbf{w}^T \tilde{\mathbf{x}}_1(n-1))^2 \mid |x(n) \right. \\ &\quad \left. - \rho \mathbf{w}^T \tilde{\mathbf{x}}_1(n-1)| < \Delta_2, |z(n-1)| < \Delta_1 \right], \\ &= \mathbb{E} \left[(\rho z(n-1) + z(n))^2 \mid |\rho z(n-1) \right. \\ &\quad \left. + z(n)| < \Delta_2, |z(n-1)| < \Delta_1 \right]. \end{aligned} \quad (42)$$

The expectation in (42) can be computed as

$$\begin{aligned} \text{MSE}_2^{\text{CDE-PD}} &= \int \int_{-\infty}^{\infty} (z(n) + \rho z(n-1))^2 f(z(n) + \rho z(n-1) \mid \\ &\quad |z(n-1)| < \Delta_1, |\rho z(n-1) \\ &\quad + z(n)| < \Delta_2) dz(n) dz(n-1). \end{aligned} \quad (43)$$

This expression is actually the computation of the variance of a bivariate truncated normal distribution. The solution of a singly truncated bivariate distribution can be found in [23]. For higher orders, i.e., $t > 2$, the solution refers to the calculation of the variance of a truncated multivariate normal distribution [24]. Although a solution already exists in the literature, it turns out to be quite complex. Moreover, its complexity increases in t . For that reason, we are considering the following approximation:

$$\begin{aligned} &\{\rho z(n-1) + z(n) \mid |z(n-1)| < \Delta_1\} \\ &\sim \mathcal{N}(0, \mathbb{E}[(\rho z(n-1) + z(n))^2 \mid |z(n-1)| < \Delta_1]), \end{aligned} \quad (44)$$

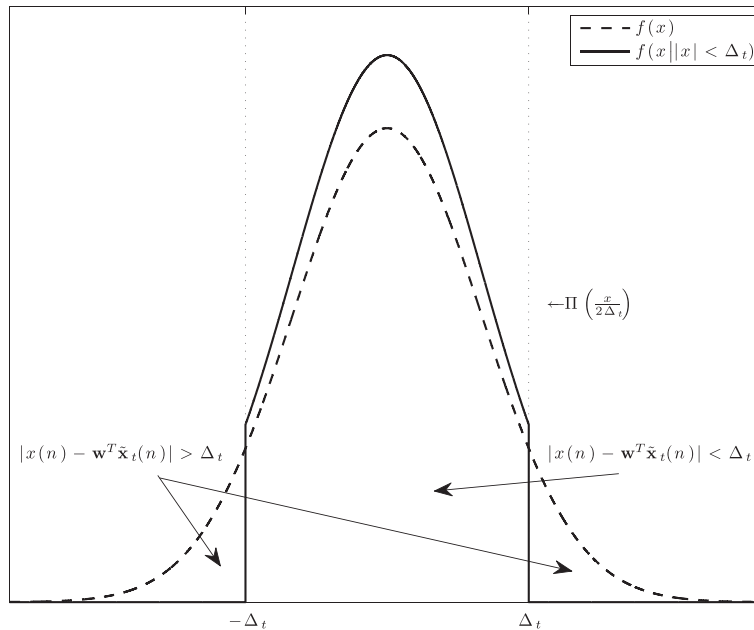


Figure 3 Qualitative representation of the conditional pdf $f(x) \mid |x| < \Delta_t$. Due to the measurements that $|x_s(n) - \mathbf{w}^T \tilde{\mathbf{x}}_t(n)| < \Delta_t$ are not introducing error since they are not estimated. The parameter Δ_t should be chosen in order to guarantee that a fraction γ of the total measurements is transmitted.

but in the general case, it does not necessarily follow a Gaussian distribution. The variance $\mathbb{E}[(\rho z(n-1) + z(n))^2 | |z(n-1)| < \Delta_1]$ can also be expressed as

$$\begin{aligned} & \mathbb{E}[(\rho z(n-1) + z(n))^2 | |z(n-1)| < \Delta_1] \\ &= \mathbb{E}[z(n)] + \rho^2 \mathbb{E}[z(n-1)^2 | |z(n-1)| < \Delta_1], \\ &= 1 - \rho^2 + \rho^2 \text{MSE}_1^{\text{CDE-PD}}, \end{aligned} \quad (45)$$

so, the MSE introduced at $t = 2$ is approximated by

$$\text{MSE}_2^{\text{CDE-PD}} \simeq h(1 - \rho^2 + \rho^2 \text{MSE}_1^{\text{CDE-PD}} | \Delta_2). \quad (46)$$

It is easy to conclude that for the general case t , the $\text{MSE}_t^{\text{CDE-PD}}$ is

$$\text{MSE}_t^{\text{CDE-PD}} \simeq \text{MSE}_t^{\text{CDE-PD}} = h(1 - \rho^2 + \rho^2 \text{MSE}_{t-1}^{\text{CDE-PD}} | \Delta_t). \quad (47)$$

□

Hence, the $\mathcal{D}(\text{CDE}, \text{PD})$ is approximated by

$$\mathcal{D}(\text{CDE}, \text{PD}) \simeq \sum_{t=0}^{\infty} P_t \text{MSE}_t^{\text{CDE-PD}}. \quad (48)$$

However, this is still an open problem. It is because the values of P_t are not determined yet. We study this issue afterwards in Section 4.5.

4.4.2 The pair CDE-SD

If $\hat{x}(n)$ is constructed from a linear prediction using the LWF, the MSE in prediction is directly $\sigma_z^2 = 1 - \rho^2$. However, using other strategies, the error will increase as we have seen in (18). In particular, the pair CDE-SD constructs $\hat{x}(n)$ as the last transmitted sample, i.e., $\hat{x}(n) = x(n-t)$. This prediction scheme introduces an error not only due to $z(n)$ but also due to $x(n)$.

Lemma 3. The $\text{MSE}_t^{\text{CDE-SD}}$ is an approximation of $\text{MSE}_t^{\text{CDE-SD}}$ (i.e., the error introduced by the CDE-SD pair at the state t) and it is defined as

$$\text{MSE}_t^{\text{CDE-SD}} = h(1 - \rho^2 + \text{MSE}_{t-1}^{\text{CDE-SD}} | \Delta_t) \leq \text{MSE}_t^{\text{CDE-SD}}. \quad (49)$$

Proof. Similarly to the CDE-SD, for $t = 1$ the error $\text{MSE}_1^{\text{CDE-SD}}$ follows the conditional variance such that

$$\begin{aligned} \text{MSE}_1^{\text{CDE-SD}} &= \mathbb{E}[(x(n) - x(n-1))^2 | |x(n) - x(n-1)| < \Delta_1] \\ &= \mathbb{E}[(z(n) - (1 - \rho)x(n-1))^2 | |z(n) - (1 - \rho)x(n-1)| < \Delta_1] \\ &= \mathbb{E}[z'(n)^2 | |z'(n)| < \Delta_1] \\ &= \int_{-\infty}^{\infty} z'(n)^2 f(z'(n) | |z'(n)| < \Delta_1) dz'(n), \end{aligned} \quad (50)$$

where $z'(n) = z(n) - (1 - \rho)x(n-1)$ contains both the error contribution due to $z(n)$ and $x(n)$ with the variance $\sigma_z'^2$ equal to

$$\begin{aligned} \sigma_z'^2 &= \mathbb{E}[(z(n) - (1 - \rho)x(n-1))^2] \mathbb{E}[z(n)] \\ &\quad + (1 - \rho)^2 \mathbb{E}[x(n-1)] = 2(1 - \rho). \end{aligned} \quad (51)$$

Therefore, the $\text{MSE}_1^{\text{CDE-SD}}$ is

$$\text{MSE}_1^{\text{CDE-SD}} = h(2(1 - \rho) | \Delta_1). \quad (52)$$

For $t = 2$ the available information is twofold: (1) we know that $|x(n) - x(n-2)| < \Delta_2$, and (2) we also know that in $t = 1$ the error was $|z'(n-1)| < \Delta_1$. Therefore, the $\text{MSE}_2^{\text{CDE-SD}}$ can be written as

$$\begin{aligned} \text{MSE}_2^{\text{CDE-SD}} &= \mathbb{E}[(x(n) - x(n-2))^2 | |x(n) - x(n-2)| < \Delta_2, |z'(n-1)| < \Delta_1] \\ &= \mathbb{E}[(\rho z(n-1) + z(n) - (1 - \rho^2)x(n-2))^2 | \\ &\quad | \rho z(n-1) + z(n) - (1 - \rho^2)x(n-2) | \\ &\quad < \Delta_2, |z'(n-1)| < \Delta_1]. \end{aligned} \quad (53)$$

To solve the $\text{MSE}_t^{\text{CDE-SD}}$ in a recursive way may be harder than for the CDE-PD case. It is because we cannot apply directly the conditional function since the expectation in (53) is not of the form $h(\sigma_x^2 | \Delta) = \mathbb{E}[x^2 | |x| < \Delta]$. Hence, to simplify, we propose a lower bound for (53) such that

$$\begin{aligned} \text{MSE}_2^{\text{CDE-SD}} &\geq \mathbb{E}[(z'(n-1) + z(n))^2 | |z'(n-1) + z(n)| < \Delta_2, |z'(n-1)| < \Delta_1]. \end{aligned} \quad (54)$$

One can easily check that it is in fact a lower bound since

$$\begin{aligned} \mathbb{E}[(z(n) - (1 - \rho)x(n-1))^2] &\leq \mathbb{E}[(\rho z(n) - (1 - \rho^2)x(n-1))^2] \\ (1 - \rho^2) &\leq 2(1 - \rho). \end{aligned} \quad (55)$$

Our proposed lower bound is very close to the real value for high values of ρ . Using the same approximation as in the CDE-PD case, and after some simple algebra, we can find the lower bound of (53) as

$$\text{MSE}_2^{\text{CDE-SD}} = h(1 - \rho^2 + \text{MSE}_1^{\text{CDE-SD}} | \Delta_2) \leq \text{MSE}_2^{\text{CDE-SD}}. \quad (56)$$

It is easy to conclude that for the general case t , the $\text{MSE}_t^{\text{CDE-SD}}$ is

$$\text{MSE}_t^{\text{CDE-SD}} = h(1 - \rho^2 + \text{MSE}_{t-1}^{\text{CDE-SD}} | \Delta_t) \leq \text{MSE}_t^{\text{CDE-SD}}. \quad (57)$$

□

Hence, the $\mathcal{D}(\text{CDE}, \text{SD})$ is lower-bounded by

$$\mathcal{D}(\text{CDE}, \text{SD}) \geq \sum_{t=0}^{\infty} P_t \text{MSE}_t^{\text{CDE-SD}}. \quad (58)$$

As for the case of the CDE-PD pair, this is still an open problem, and it is studied afterwards in Section 4.5.

4.5 Design of the CDE-SD and the CDE-PD

From the design point of view, our aim is to obtain a set of Δ_t 's that assure a coding rate at the CDE of γ . However, there are infinite solutions as we pointed out in (24). That is why we propose two possible approaches to face with the design of Δ_t :

- Fixed Δ_t , i.e., $\Delta_t = \Delta$ for all t .
- Variable Δ_t in order to maintain constant transition probabilities, i.e., $p_{t-1,t} = p$ for all t .

4.5.1 Fixed Δ_t design

This is probably the simplest approach to design the CDE since the encoder does not have to change the value of Δ_t according to the current state since $\Delta_t = \Delta$ for all t .

First, we want to make explicit the existing relation between Δ and $p_{t-1,t}$, as

$$p_{t-1,t}(\Delta) = \int_{-\Delta}^{\Delta} f_t(x) dx, \quad (59)$$

where $f_t(x)$ is the pdf of the error at state t .

Following the assumption in (44), the variable $x(n) - \hat{x}(n)$ follows a Gaussian distribution with zero mean and variance $\text{MSE}_t^{\text{CDE}}(\Delta)$, where

$$\text{MSE}_t^{\text{CDE}}(\Delta) = \begin{cases} 1 - \rho^2 + \text{MSE}_t^{\text{CDE-SD}}(\Delta) & \text{if CDE-SD} \\ 1 - \rho^2 + \rho^2 \text{MSE}_t^{\text{CDE-PD}}(\Delta) & \text{if CDE-PD.} \end{cases} \quad (60)$$

Thus^c,

$$p_{t-1,t}(\Delta) = 1 - 2 \int_{\Delta}^{\infty} f_t(x) dx = \text{erf} \left(\frac{\Delta}{\sqrt{2 \text{MSE}_{t-1}^{\text{CDE}}(\Delta)}} \right), \quad (61)$$

where $\text{erf}(x)$ is the error function of x . Using the result in (24), we can numerically approximate Δ that assures $P_0 = \gamma$ as the unique solution of

$$\sum_{i=1}^T \left(\prod_{t=1}^i \text{erf} \left(\frac{\Delta}{\sqrt{2 \text{MSE}_{t-1}^{\text{CDE}}(\Delta)}} \right) \right) = \frac{1-\gamma}{\gamma}, \text{ for } T \rightarrow \infty. \quad (62)$$

The solution of Δ for the different values of γ and ρ can be graphically seen in Figure 4.

4.5.2 Variable Δ_t design

This approach allows for a slightly easier computation of the values of Δ_t . The main difference with the previous design scheme is that we can use the result in the following lemma:

Lemma 4. *The uniform solutions of the non-zero transition probabilities and for the stationary probability vector are*

$$p_{0,0} = \gamma; p_{t-1,t} = 1 - \gamma, \text{ for } t = 1, 2, \dots; P_t = \gamma(1 - \gamma)^t. \quad (63)$$

Proof. Let us first impose that $P_0 = \gamma$. Hence, for the uniform probability case $p_{t-1,t} = p$, and using (24)

$$\frac{1-\gamma}{\gamma} = \sum_{t=1}^{\infty} p^t, \quad \frac{1-\gamma}{\gamma} - 1 = \frac{1}{1-p}, \quad p = 1 - \gamma. \quad (64)$$

So, if $p_{0,1} = 1 - \gamma$, we obtain that $p_{0,0} = \gamma$. In order to compute the probability of each state, and considering (23), we get

$$P_t = \gamma p^t = \gamma(1 - \gamma)^t. \quad (65)$$

□

Hence, Δ_t is directly

$$\Delta_t = \sqrt{2 \text{MSE}_{t-1}^{\text{CDE}}(\Delta_{t-1})} \text{erf}^{-1}(1 - \gamma), \text{ for } t = 1, 2, \dots, \quad (66)$$

where $\text{MSE}_0^{\text{CDE-}\{\text{SD}, \text{PD}\}} = 0$; hence, $\text{MSE}_0^{\text{CDE}} = 1 - \rho^2$ (as in (60)).

To graphically validate our design framework, we have proposed the following experiment:

Experiment 1. *We have simulated the CDE-SD and the CDE-PD for $\gamma = [1/8 \ 1/4 \ 1/2]$ and for $\rho \in [0, 1]$. The signal has been generated following the AR-1 process of 5,000 samples (for each value of ρ). We have computed the probability of transmission P_0 obtained using our threshold design framework.*

From Experiment 1, we have plotted the probability of transmission P_0 as a function of ρ and for each value of γ . We have used the variable Δ_t design. In Figure 5, we have compared the obtained results with the target coding rate γ , and we have observed that for the case of CDE-PD, the fitting is very accurate. For the case of CDE-SD, it is slightly worse. It is due to the approximation in (53). However, we have said that this approximation improves for $\rho \rightarrow 1$. This behavior can be observed in Figure 5.

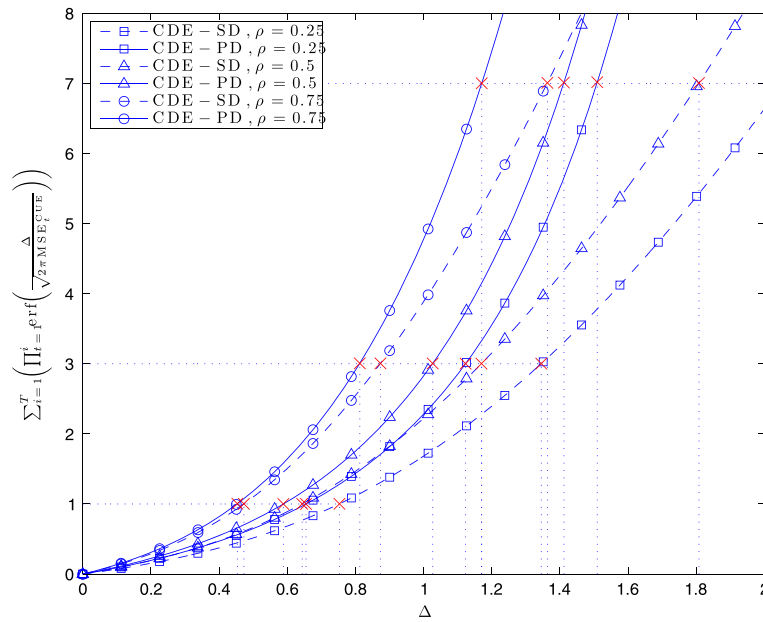


Figure 4 Numerical solution of Δ for the fixed Δ_t design. The values of γ are $\gamma = \{0.125, 0.25, 0.5\}$, the values of ρ are $\rho = \{0.25, 0.5, 0.75\}$, and the value of T is 100. The red cross marks are the Δ solutions for a given ρ and γ .

5 Downsampling distortion of other typical strategies

In order to measure the performance of the CDE, we also evaluate the performance of different encoder-decoder pairs in terms of the downsampling distortion. These are DDE-SD, DDE-PD, PDE-SD, and PDE-PD.

5.1 The pair DDE-SD

The index t denotes the time spacing between the last available sample with the current one. Thus, we can compute the $\text{MSE}_t^{\text{DDE-SD}}$ using the result in (18) for each observation vector $\tilde{\mathbf{x}}_t$. Therefore, the downsampling distortion will be the sum of the MSE contributions for each

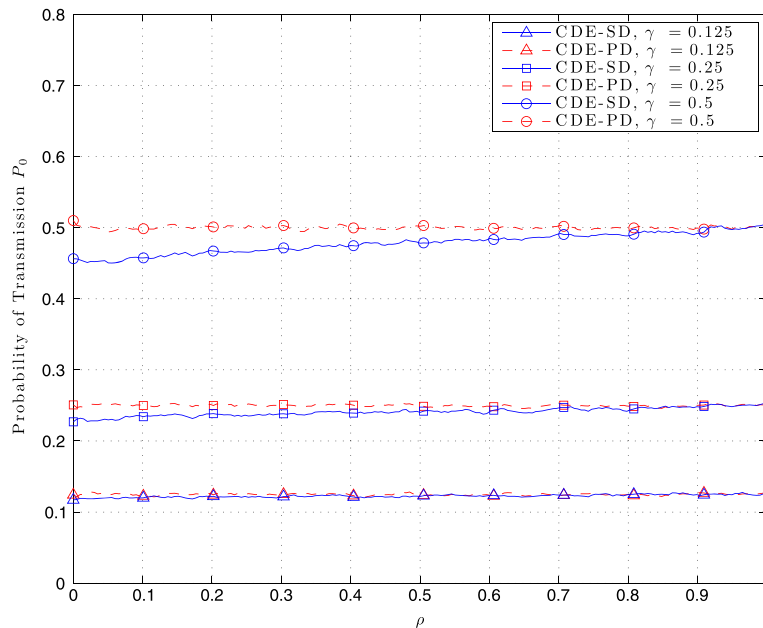


Figure 5 Experimental results from Experiment 1. The empirical probability of transmission is compared with the target coding rate γ for the CDE-SD and CDE-PD schemes. We have used the variable threshold design.

state. Applying the definition of stationary probability vector in Definition 5, we extract that $P_i = P_j$ for all $i, j = 0, 1, \dots, T$. Since we impose a coding rate of γ , the probability of transmission, i.e., P_0 , is $P_0 = 1/T = \gamma$. The stationary probability vector is $\mathbf{p} = \gamma \mathbf{1}$. Hence, the downsampling distortion for the DDE-SD can be computed as

$$\begin{aligned} \mathcal{D}(\text{DDE}, \text{SD}) &= \sum_{t=0}^{T-1} P_t \text{MSE}_t^{\text{DDE-SD}} = \frac{2}{T} \sum_{t=0}^{T-1} (1 - \rho^t) = 2 - \frac{2}{T} \sum_{t=0}^{T-1} \rho^t \\ &= 2 - \frac{2}{T} \frac{\rho^{T-1}}{\rho - 1} = 2 - 2\gamma \frac{\rho^{1/\gamma} - 1}{\rho - 1}. \end{aligned} \quad (67)$$

5.2 The pair DDE-PD

The knowledge of the correlation parameters is available at the PD, and hence, it can predict the non-transmitted samples using the LWF. Following Theorem 1, the $\text{MSE}_t^{\text{DDE-PD}} = 1 - \rho^{2t}$. Hence, the downsampling distortion for the DDE-PD can be computed as

$$\begin{aligned} \mathcal{D}(\text{DDE}, \text{PD}) &= \sum_{t=0}^{T-1} P_t \text{MSE}_t^{\text{DDE-PD}} = \frac{1}{T} \sum_{t=0}^{T-1} (1 - \rho^{2t}) \\ &= 1 - \frac{1}{T} \sum_{t=0}^{T-1} \rho^{2t} \\ &= 1 - \frac{1}{T} \frac{\rho^{2T} - 1}{\rho^2 - 1} = 1 - \gamma \frac{\rho^{2/\gamma} - 1}{\rho^2 - 1}. \end{aligned} \quad (68)$$

5.3 The pair PDE-SD

The PDE can also be modeled following the infinite MC in Figure 2. Hence, the transmission matrix \mathbf{T}_{CDE} has the same structure than \mathbf{T}_{CDE} in (22), and the expressions (23) and (24) are valid as well. However, the rest is different.

For simplicity, we assume that all $p_{t-1,t}$ are equal, i.e., the *uniform probability* case. The results of Lemma 4 also apply here. It gives us two main advantages:

1. It is the easiest solution to be implemented in practice. The source decides either to transmit or not regardless of what the current state t is.
2. It reduces the problem to a closed-form solution.

Using the results for the MSE_t in (18) corresponding to the decoder SD, we obtain

$$\begin{aligned} \mathcal{D}(\text{PDE}, \text{SD}) &= \sum_{t=0}^{\infty} P_t \text{MSE}_t^{\text{PDE-SD}} \\ &= \sum_{t=0}^{\infty} \gamma (1 - \gamma)^t 2 (1 - \rho^t) \\ &= 2\gamma \sum_{t=0}^{\infty} ((1 - \gamma)^t - \rho^t (1 - \gamma)^t) \\ &= 2\gamma \left(\frac{1}{1 - (1 - \gamma)} - \frac{1}{1 - \rho(1 - \gamma)} \right) \\ &= 2 \left(1 - \frac{\gamma}{1 - \rho(1 - \gamma)} \right). \end{aligned} \quad (69)$$

5.4 The pair PDE-PD

The MSE associated to the state t obeys Theorem 1. The downsampling distortion for the PDE-PD can be computed as

$$\begin{aligned} \mathcal{D}(\text{PDE}, \text{PD}) &= \sum_{t=0}^{\infty} P_t \text{MSE}_t^{\text{PDE-PD}} \\ &= \sum_{t=0}^{\infty} \gamma (1 - \gamma)^t (1 - \rho^{2t}) \\ &= \gamma \sum_{t=0}^{\infty} ((1 - \gamma)^t - \rho^{2t} (1 - \gamma)^t) \\ &= \gamma \left(\frac{1}{1 - (1 - \gamma)} - \frac{1}{1 - \rho^2(1 - \gamma)} \right) \\ &= 1 - \frac{\gamma}{1 - \rho^2(1 - \gamma)}. \end{aligned} \quad (70)$$

6 Performance evaluation

In this section, we evaluate and compare the performance of the different encoder-decoder pairs as a function of the downsampling distortion. Moreover, we introduce an experimental evaluation in order to confirm the validity of our theoretical results. For that, we have generated a signal $x(n)$ as a sequence of 5,000 samples using the AR-1 model in (2) and for different values of the autoregressive parameter $\rho \in [0, 1]$ with resolution 0.01. The results are computed for $\gamma = [1/8, 1/4, 1/2]$.

6.1 The pair DDE-SD and the pair DDE-PD

We analyze the downsampling distortion for the DDE-SD and DDE-PD pairs. We compare the theoretical results with the experimental results. So, Figure 6 confirms the validity of our theoretical model for the downsampling distortion.

Also, we compare the difference in performance according to the decoder used. The PD takes into account the signal correlation information in the decoding process, and hence, the total performance is increased notably for low values of ρ . On the contrary, if $\rho \rightarrow 1$, both decoders perform similarly since $x(n) - \rho^t x(n-t) \approx x(n) - x(n-t)$.

In Figure 6, we can also graphically evaluate the impact of γ . In our scenario, the signal $x(n)$ is transmitted by the DDE in $\{8, 4, 2\}$ times following a uniform pattern. It is easy to see that the larger the γ , the lower is the distortion. However, there exists a trade-off between the downsampling distortion and the compression rate.

6.2 The pair PDE-SD and the pair PDE-PD

The downsampling distortion for the PDE-SD and the PDE-PD is plotted in Figure 7. However, the conclusions that can be extracted from these results with respect to the accuracy of the proposed analytical model and the behavior of ρ and γ with respect to the downsampling distortion

are similar to the ones established in Section 6.1. For the sake of clarity, we compare the downsampling distortion results of the different pairs later in Section 6.4.

6.3 The pair CDE-SD and the pair CDE-PD

The performance of the previous encoder-decoder pairs can be notably improved by conditional transmission at the encoder site. In particular, we study and compare the downsampling distortion of the two design approaches, i.e., the fixed Δ_t design and the variable Δ_t design (with uniform transition probabilities), depicted in Figures 8 and 9, respectively. As in the previous pairs, we compare both the experimental results with the theoretical results. However, in that case, our theoretical results are limited to an approximation rather than the real system performance. Even so, we can observe that the approximations are very accurate for all the different simulations. For the case of CDE-PD, the approximation is so close to the system performance that the difference cannot be observed because it is masked by the small amount of noise due to the simulation. For the case of CDE-SD, the difference is slightly bigger because of the approximation in (55).

Another conclusion is that the downsampling distortion is notably higher for the fixed design. It is because their transition probabilities $p_{t-1,t}$ are increasing in t , and it facilitates to achieve higher states t in the MC with higher probability (i.e., higher MSE_t 's). On the contrary, the variable design concentrates the states in lower t values.

From a practical point of view, the CDE is simpler if it follows a fixed design since the encoder only needs to know the value of Δ and also it does not need to track the current state t . However, from a computational point of view, the variable approach is simpler since it can be computed analytically, instead of numerically.

6.4 Comparison of the downsampling distortion

Finally, we compare the performance of the different encoder-decoder pairs. Although Figure 10 does not provide any extra information, it allows us to better compare the performance of the different schemes. For the sake of simplicity, we only compare the theoretical results for the case of $\gamma = 0.25$.

It can be observed that the performance of the DDE and PDE are similar. However, the deterministic encoder works slightly better since it only uses the lowest γ^{-1} states of the finite MC while PDE uses higher states that are related to higher errors. However, the main disadvantage of the DDE in front of the PDE is its lack of flexibility since the uniform solution is only valid for natural values of γ^{-1} . Furthermore, the PDE with uniform transition probabilities does not need to track the current state t of the process, and hence, it is simpler.

The big hop in performance is observed for the CDE. This encoder eliminates the transmissions of the samples with the most redundant information. Thus, only the most 'unpredictable' samples are transmitted.

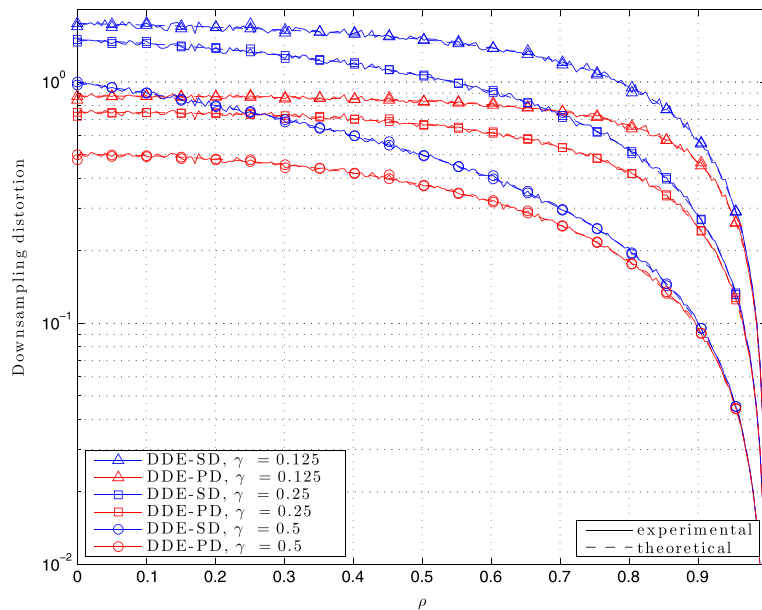


Figure 6 Experimental and theoretical downsampling distortion of the pairs DDE-SD and DDE-PD as a function of ρ . The coding rates are $\gamma = \{0.125, 0.25, 0.5\}$. Different markers denote different values of γ ; curves in blue represent the SD, whereas curves in red represent the PD; solid lines represent experimental results, and dashed lines represent theoretical results.

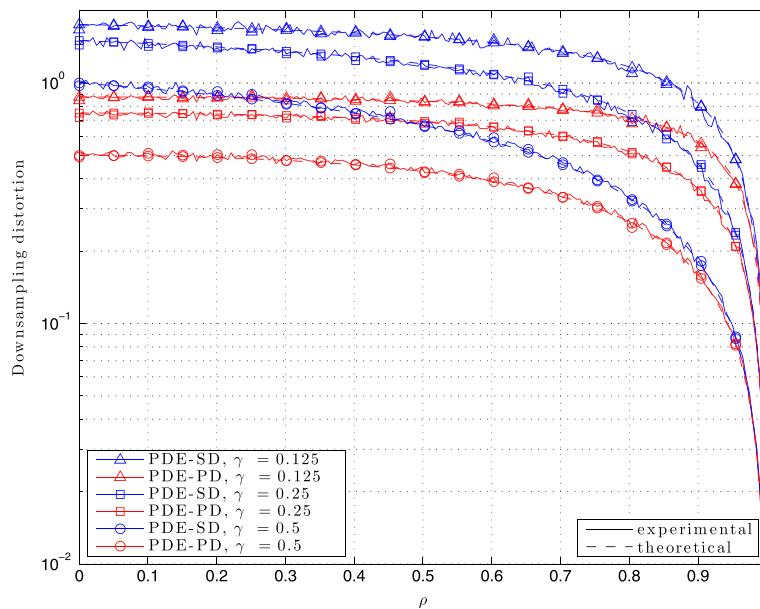


Figure 7 Experimental and theoretical downsampling distortion of the pairs PDE-SD and PDE-PD as a function of ρ . The coding rates are $\gamma = \{0.125, 0.25, 0.5\}$. Different markers denote different values of γ ; curves in blue represent the SD, whereas curves in red represent the PD; solid lines represent experimental results, and dashed lines represent theoretical results.

7 Conclusions

In this chapter, we have evaluated the performance of different encoding-decoding strategies in order to reduce the number of transmitted samples and hence to decrease the power spent in transmission. We have presented them as

an energy-efficient solution for the wireless sensor network communication problem. In particular, we define the downsampling distortion function in order to evaluate the performance in terms of the trade-off between compression rate and distortion at the fusion center of the

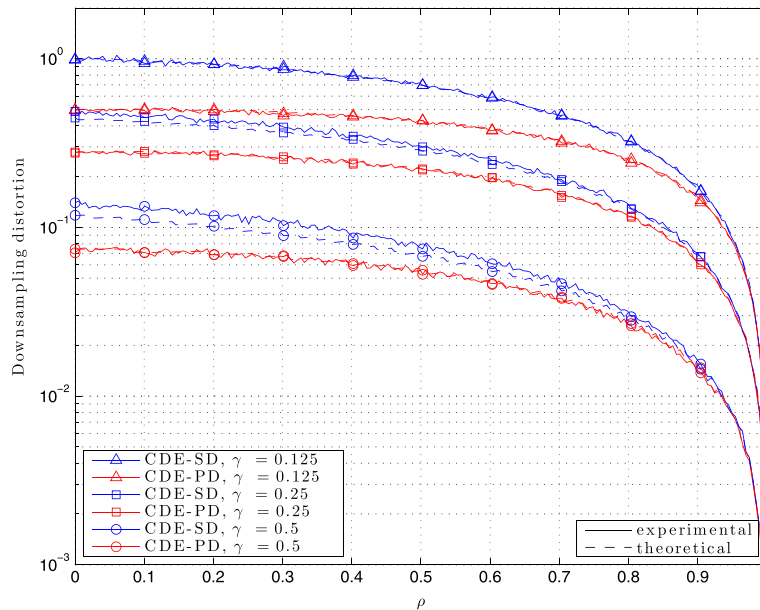


Figure 8 Experimental and theoretical approximation of downsampling distortion of CDE-SD/CDE-PD pairs following a fixed Δ_t design. The coding rates are $\gamma = \{0.125, 0.25, 0.5\}$. Different markers denote different values of γ ; curves in blue represent the SD, whereas curves in red represent the PD; solid lines represent experimental results, and dashed lines represent theoretical results.

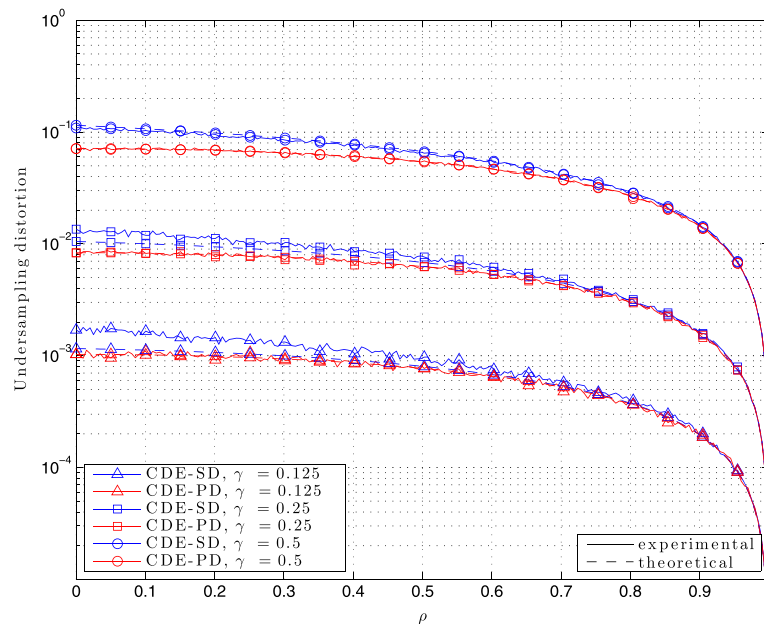


Figure 9 Experimental and theoretical approximation of downsampling distortion of CDE-SD/CDE-PD pairs following a variable Δ_t design. The coding rates are $\gamma = \{0.125, 0.25, 0.5\}$. Different markers denote different values of γ ; curves in blue represent the SD, whereas curves in red represent the PD; solid lines represent experimental results, and dashed lines represent theoretical results.

combination of three downsampling encoders, which are the DDE, the PDE, and the CDE, with two decoders: the SD and the PD.

We have obtained closed-form expressions for the pairs DDE-SD, DDE-PD, PDE-SD, and PDE-PD and accurate approximations for CDE-SD and CDE-PD. Moreover, we

have proposed two strategies in order to design the threshold of the condition in the CDE, i.e., the fixed threshold design and the variable threshold design.

The simulation results validate our theoretical results. Furthermore, we have compared the performance of the different pairs and showed the impact of taking into

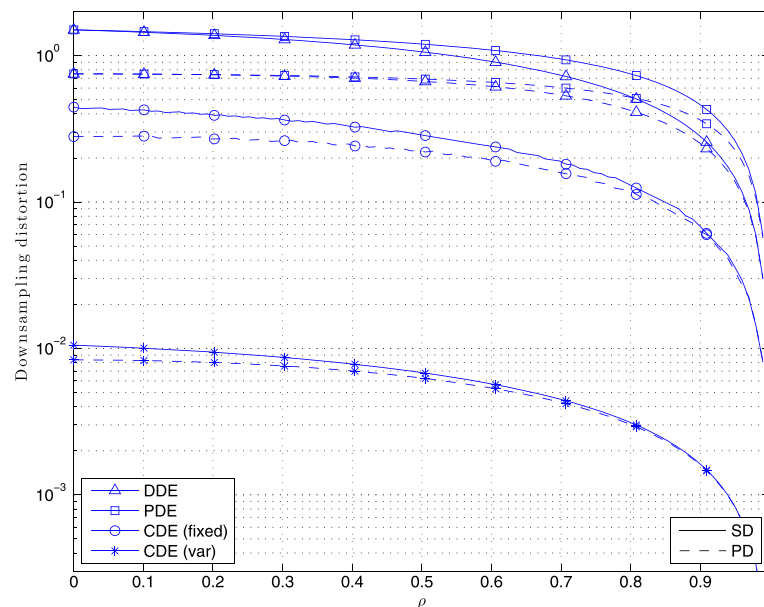


Figure 10 Comparison of the downsampling distortion of the different encoding-decoding pairs as a function of ρ . The coding rate is $\gamma = 0.25$.

account the signal model in the encoding-decoding process. Hence, the pair CDE-PD (with variable threshold design) outperforms by far the rest of the studied strategies. However, extending the CDE analysis for higher order AR models or even for other time-correlated signal models remains as an open problem.

Endnotes

^a **Notation.** Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors, and italics denote scalars. $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ denote transpose, complex conjugate, and conjugate transpose (Hermitian), respectively. $[X]_{i,j}$ and $[x]_i$ are the $(i$ th, j th) element of matrix X and the i th position of vector x , respectively. $[X]_i$ denotes the i th column of X . $|\cdot|$ is the absolute value. $\|a\|$ represents the Euclidean norm of a . Let \hat{a} refer to the estimated value of variable a . $\mathbb{E}[\cdot]$ is the statistical expectation. Function $\text{erf}(\cdot)$ represents the error function.

^b The **conditional variance** of a continuous random variable X given the condition $Y = y$ is defined as $\text{var}(X|Y = y) = \mathbb{E}[X^2|Y = y] - \mathbb{E}[X|Y = y]^2$, where $f(X|Y = y)$ is the conditional pdf of X given $Y = y$.

^c It comes from the definition of the cumulative density function of a Gaussian variable such that $\int_{-\infty}^a f(x)dx = \frac{1}{2} \left(1 + \text{erf} \left(\frac{a}{\sqrt{2}\sigma_a} \right) \right)$.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work is supported by the Spanish Government under project TEC2011-28219 and the Catalan Government under grant 2009 SGR 298.

Received: 14 January 2013 Accepted: 30 April 2013

Published: 10 May 2013

References

1. CK Toh, Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks. *Commun. Mag., IEEE*. **39**(6), 138–147 (2001)
2. O Younis, S Fahmy, HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *Mobile Comput. IEEE Trans.* **3**(4), 366–379 (2004)
3. R Mudumbai, D Brown, U Madhow, H Poor, Distributed transmit beamforming: challenges and recent progress. *Commun. Mag., IEEE*. **47**(2), 102–110 (2009)
4. K Zarifi, S Zaidi, S Affes, A Ghayeb, A distributed amplify-and-forward beamforming technique in wireless sensor networks. *Signal Process., IEEE Trans.* **59**(8), 3657–3674 (2011)
5. S Pradhan, J Kusuma, K Ramchandran, Distributed compression in a dense microsensor network. *Signal Process. Mag., IEEE*. **19**(2), 51–60 (2002)
6. N Sun, J Wu, Optimum sampling in spatial-temporally correlated wireless sensor networks. *EURASIP J. Wireless Commun. Netw.* **2013**, 5 (2013)
7. D Neuhoff, R Gilbert, Causal source codes. *Inf. Theory, IEEE Trans.* **28**(5), 701–713 (1982)
8. T Weissman, N Merhav, On causal source codes with side information. *Inf. Theory, IEEE Trans.* **51**(11), 4003–4013 (2005)
9. M Derpich, Improved upper bounds to the causal quadratic rate-distortion function for Gaussian stationary. *Inf. Theory, IEEE Trans.* **58**(99), 3131–3152 (2012)
10. H Viswanathan, T Berger, Sequential coding of correlated sources. *Inf. Theory, IEEE Trans.* **46**, 236–246 (2000)

11. R Zamir, Y Kochman, U Erez, Achieving the Gaussian rate distortion function by prediction. *Inf. Theory, IEEE Trans.* **54**(7), 3354–3364 (2008)
12. JB O'Neal, Delta-modulation quantizing noise-analytic and computer simulation results for Gaussian and television input signals. *Bell Syst. Tech. J.* **45**, 117–141 (1971)
13. EN Protonotarios, Slope overload noise in differential pulse code modulation systems. *Bell Syst. Tech. J.* **46**, 2119–2161 (1967)
14. TM Cover, JA Thomas, *Elements on Information Theory*. (Wiley, New York, 1991)
15. JB O'Neal, Signal-to-quantizing-noise ratio for differential PCM. *IEEE Trans. Commun. Technol.* **19**, 568–570 (1971)
16. N Farvardin, J Modestino, Rate-distortion performance of DPCM schemes for autoregressive sources. *Inf. Theory, IEEE Trans.* **31**(3), 402–418 (1985)
17. O Guleryuz, M Orchard, On the DPCM compression of Gaussian autoregressive sequences. *Inf. Theory, IEEE Trans.* **47**(3), 945–956 (2001)
18. R Rugin, A Conti, G Mazzini, in *Proceedings of the 15th International Conference on Software, Telecommunications and Computer Networks, 2007. SoftCOM 2007*. Experimental investigation of the energy consumption for wireless sensor network with centralized data collection scheme (Split-Dubrovnik, 27–29 Sept 2007), pp. 1–5
19. Q Wang, Traffic analysis, modeling and their applications in energy-constrained wireless sensor networks: on network optimization and anomaly detection. (Mid Sweden University, 2010). <http://urn.kb.se/resolve?urn=urn:nbn:se:miun:diva-10690>. Accessed 15 July 2012
20. T Hashimoto, S Arimoto, On the rate-distortion function for the nonstationary Gaussian autoregressive process. *Inf. Theory, IEEE Trans.* **26**(4), 478–480 (1980)
21. S Haykin, *Adaptive Filter Theory*, 4th edn. (Prentice Hall, Upper Saddle River, 2001)
22. J Barceló-Lladó, A Morell, G Seco-Granados, Enhanced correlation estimators for distributed source coding in large wireless sensor networks. *IEEE Sensors J.* **12**(9), 2799–2806 (2012)
23. S Rosenbaum, Moments of a truncated bivariate normal distribution. *J. R. Stat. Soc. Ser. B (Methodological)*. **23**(2), 405–408 (1961)
24. BG Manjunath, Wilhelm S, *Moments calculation for the double truncated multivariate normal density (Social Science Research Network 2009)*. <http://dx.doi.org/10.2139/ssrn.1472153>. Accessed 20 Aug 2012

doi:10.1186/1687-6180-2013-101

Cite this article as: Barceló-Lladó et al.: Conditional downsampling for energy-efficient communications in wireless sensor networks. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:101.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com